K-Nearest Neighbors: The Simplest ML Algorithm

Nipun Batra

IIT Gandhinagar

July 30, 2025

Outline

The KNN Intuition

Core idea: "You are the average of your closest friends"

Real-life Example

To predict house price, look at prices of similar houses nearby:

The KNN Intuition

Core idea: "You are the average of your closest friends"

Real-life Example

To predict house price, look at prices of similar houses nearby:

- Same neighborhood

The KNN Intuition

Core idea: "You are the average of your closest friends"

Real-life Example

To predict house price, look at prices of similar houses nearby:

- Same neighborhood
- Similar size

The KNN Intuition

Core idea: "You are the average of your closest friends"

Real-life Example

To predict house price, look at prices of similar houses nearby:

- Same neighborhood
- Similar size
- Similar age

The KNN Intuition

Core idea: "You are the average of your closest friends"

Real-life Example

To predict house price, look at prices of similar houses nearby:

- Same neighborhood
- Similar size
- Similar age

The KNN Intuition

Core idea: "You are the average of your closest friends"

Real-life Example

To predict house price, look at prices of similar houses nearby:

- Same neighborhood
- Similar size
- Similar age

KNN Algorithm

The KNN Intuition

Core idea: "You are the average of your closest friends"

Real-life Example

To predict house price, look at prices of similar houses nearby:

- Same neighborhood
- Similar size
- Similar age

KNN Algorithm

1. Find k nearest neighbors to query point

The KNN Intuition

Core idea: "You are the average of your closest friends"

Real-life Example

To predict house price, look at prices of similar houses nearby:

- Same neighborhood
- Similar size
- Similar age

KNN Algorithm

1. Find *k* nearest neighbors to query point
2. For **classification**: Vote (majority class)

The KNN Intuition

Core idea: "You are the average of your closest friends"

Real-life Example

To predict house price, look at prices of similar houses nearby:

- Same neighborhood
- Similar size
- Similar age

KNN Algorithm

1. Find k nearest neighbors to query point
2. For **classification**: Vote (majority class)
3. For **regression**: Average their values

Pop Quiz: KNN Basics

Quick Quiz 1

What does KNN stand for and what does "k" represent?

a) K-Neural Networks, k = number of layers

Answer: b) K-Nearest Neighbors - k is the number of closest points we use for prediction!

Pop Quiz: KNN Basics

Quick Quiz 1

What does KNN stand for and what does "k" represent?

a) K-Neural Networks, k = number of layers
b) K-Nearest Neighbors, k = number of neighbors to consider

Answer: b) K-Nearest Neighbors - k is the number of closest points we use for prediction!

Pop Quiz: KNN Basics

Quick Quiz 1

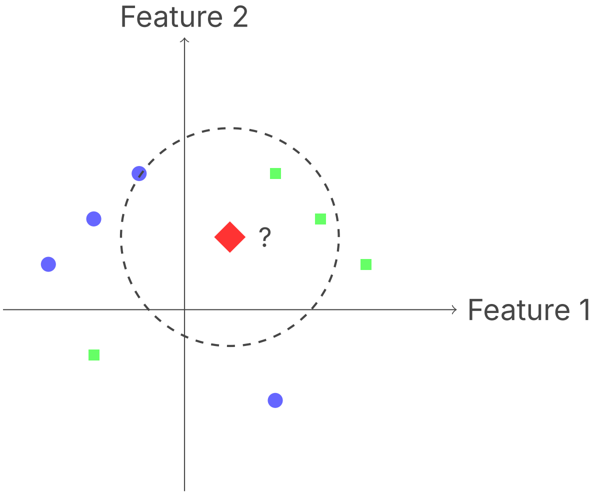
What does KNN stand for and what does "k" represent?

a) K-Neural Networks, k = number of layers
b) K-Nearest Neighbors, k = number of neighbors to consider
c) K-Naive Nets, k = number of features

Answer: b) K-Nearest Neighbors - k is the number of closest points we use for prediction!

KNN for Classification

Example: Classify new point (red ?) with k=5



Within circle: 3 blue circles, 2 green squares → Predict Class A!

Choosing the Right k

Choosing the Right k

Choosing the Right k

Small k (e.g., k=1)

Choosing the Right k

Small k (e.g., k=1)

- High variance

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk
- Complex decision boundaries

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk
- Complex decision boundaries

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk
- Complex decision boundaries

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk
- Complex decision boundaries

Large k

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk
- Complex decision boundaries

Large k

- Low variance

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk
- Complex decision boundaries

Large k

- Low variance
- High bias

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk
- Complex decision boundaries

Large k

- Low variance
- High bias
- Smoother boundaries

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk
- Complex decision boundaries

Large k

- Low variance
- High bias
- Smoother boundaries
- Underfitting risk

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk
- Complex decision boundaries

Large k

- Low variance
- High bias
- Smoother boundaries
- Underfitting risk

Choosing the Right k

Small k (e.g., k=1)

- High variance
- Sensitive to noise
- Overfitting risk
- Complex decision boundaries

Large k

- Low variance
- High bias
- Smoother boundaries
- Underfitting risk

Bias-Variance Tradeoff

Sweet spot: Use cross-validation to find optimal k (often $k = \sqrt{n}$ as starting point)

Pop Quiz: k Selection

Quick Quiz 2

For a noisy dataset, which k value would be better?

a) $k = 1$ (only nearest neighbor)

Answer: b) Larger k reduces impact of noisy points by averaging over more neighbors!

Pop Quiz: k Selection

Quick Quiz 2

For a noisy dataset, which k value would be better?

a) k = 1 (only nearest neighbor)
b) k = 15 (many neighbors)

Answer: b) Larger k reduces impact of noisy points by averaging over more neighbors!

Pop Quiz: k Selection

Quick Quiz 2

For a noisy dataset, which k value would be better?

a) k = 1 (only nearest neighbor)
b) k = 15 (many neighbors)
c) It doesn't matter

Answer: b) Larger k reduces impact of noisy points by averaging over more neighbors!

Feature Scaling is Critical!

Problem: Different feature scales dominate distance calculations

Example Without Scaling

Income differences will dominate! Age becomes irrelevant.

Feature Scaling is Critical!

Problem: Different feature scales dominate distance calculations

Example Without Scaling

- Feature 1: Age (20-80 years)

Income differences will dominate! Age becomes irrelevant.

Feature Scaling is Critical!

Problem: Different feature scales dominate distance calculations

Example Without Scaling

- Feature 1: Age (20-80 years)
- Feature 2: Income ($20,000-$200,000)

Income differences will dominate! Age becomes irrelevant.

Feature Scaling is Critical!

Problem: Different feature scales dominate distance calculations

Example Without Scaling

- Feature 1: Age (20-80 years)
- Feature 2: Income ($20,000-$200,000)

Income differences will dominate! Age becomes irrelevant.

Feature Scaling is Critical!

Problem: Different feature scales dominate distance calculations

Example Without Scaling

- Feature 1: Age (20-80 years)
- Feature 2: Income ($20,000-$200,000)

Income differences will dominate! Age becomes irrelevant.

Solution: Normalize Features

Feature Scaling is Critical!

Problem: Different feature scales dominate distance calculations

Example Without Scaling

- Feature 1: Age (20-80 years)
- Feature 2: Income ($20,000-$200,000)

Income differences will dominate! Age becomes irrelevant.

Solution: Normalize Features

- **Min-Max:** Scale to [0,1]: $x' = \frac{x - \min}{\max - \min}$

Feature Scaling is Critical!

Problem: Different feature scales dominate distance calculations

Example Without Scaling

- Feature 1: Age (20-80 years)
- Feature 2: Income ($20,000-$200,000)

Income differences will dominate! Age becomes irrelevant.

Solution: Normalize Features

- **Min-Max:** Scale to [0,1]: $x' = \frac{x - \min}{\max - \min}$
- **Z-score:** $x' = \frac{x - \mu}{\sigma}$ (mean=0, std=1)